- Introduction
- Kaplan-Meier estimator
- log-Rank test
- Hazard ratio

- survival function → Kaplan-Meier
- comparison of surv. functions → log-rank test & hazard ratio
- covariates in $H(t)$ → Cox, frailty, param. models

Outcome = survival time = time from a starting time to a particular endpoint
= time-to-event

T_i^* = true event time
C_i = time of the censoring
} $\perp\!\!\!\perp$ (non informative censoring)

T_i = survival time = $\min\{T_i^*, C_i\}$
$\delta_i = \mathbb{1}\{T_i^* \leq C_i\}$ = indicator for the non-censoring

$\Longrightarrow$ data = $\{(T_i, \delta_i)\}_i$

T = survival time
→ density function $f(t)$
→ distribution function $F(t) = \mathbb{P}(T \leq t)$
→ Survival function $S(t) = \mathbb{P}(T > t) = 1 - F(t)$
→ Hazard function $h(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t}$
= istantaneous risk of failure

T discrete

$\mathbb{P}(T = t_i) = f(t_i) = h(t_i) \cdot S(t_i)$
$S(t) = \sum_{i: t_i \geq t} f(t_i) = \prod_{i: t_i < t} (1 - h(t_i))$
$h(t_i) = 1 - \frac{S(t_{i+1})}{S(t_i)}$

T continuous

$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t)$ $(\ln(\cdot))$

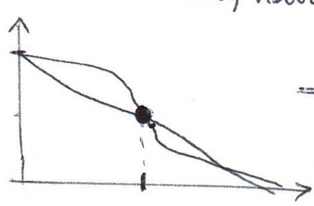
$S(t) = e^{-\int_0^t h(u) du}$ $t \geq 0$
$= e^{-H(t)}$ ($H(t) = -\ln(S(t))$ = cumulative hazard function)

$S(t)$ = estimate of the percentage of individuals who are still event-free at time t

$CFP(t) = 1 - S(t)$
= cumulative incidence / cumulative failure probability
= estimate of the % of individuals who have already experienced the event at time t

Note: if the survival function is decreasing sharply ⇒ mortality rate is high

Median survival time = time at which 50% of the individuals are still event-free
(different curves may have the same median surv. time:



⇒ medians do <u>not</u> describe the whole curve)

Estimates :

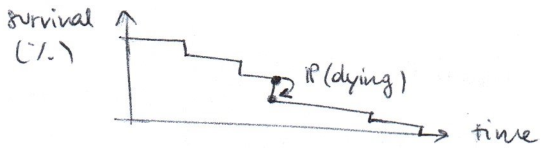
- directly estimate of the survival function
  → Kaplan-Meier estimator
- estimate of $H(t)$ (cumulative hazard function)
  → $\boxed{\text{Nelson-Aalen estimator}}$

$$\boxed{\begin{aligned} &\hat{H}(t) = \sum_{j:\, t_j^* \le t} \frac{d_j}{n_j} \\ &\hat{Var}(\hat{H}(t)) = \sum_{j:\, t_j^* \le t} \frac{d_j}{n_j^2} \\ &\text{where } n_j = \#\text{event free before } t_j^* \overset{\text{I}}{=} \#\text{ at risk at time } t_j^* \\ &\phantom{\text{where }} d_j = \#\text{ observed events at } t_j^* \end{aligned}}$$

KAPLAN-MEIER CURVE = probability of surviving in a given length of time while considering time in many small intervals

Assumptions :
1. Censoring unrelated to the outcome
2. Survival probs. are the same for subjects recruited early and late in the study
3. The events occurred at specific times

$$\hat{S}(t) = \prod_{j:\, t_j^* \le t} p_j = \prod_{j:\, t_j^* \le t} \left(1 - \frac{d_j}{n_j}\right)$$ = step function where we have jumps observed at the events times



$$\hat{h}_j = \hat{h}(t_j) = \frac{d_j}{n_j}$$ = estimate of the hazard function at time j
= # observed events at time j / # at risk at time j

$$\hat{Var}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{j:\, t_j^* \le t} \frac{d_j}{n_j(n_j - d_j)}$$

PLAIN CI :

$$CI_{0.95}(S(t)) = \left[\hat{S}(t) \pm z_{0.975}\, \hat{se}(t)\right], \qquad \hat{se}(t) = \sqrt{\hat{Var}(\hat{S}(t))}$$

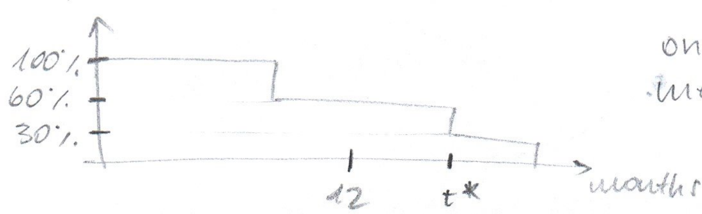
↳ produced with:

survfit (Surv(T_i, δ_i) ~ 1, data, conf.type = 'plain')

otherwise, in automatic from the survival function we get

LOG CONFIDENCE INTERVAL :

$$\underbrace{CI_{0.95}(S(t))}_{\text{more stable but more computationally demanding}} = \left[\hat{S}(t)\, e^{-z_{0.975}\sqrt{\hat{Var}(\ln(\hat{S}(t)))}} ;\ \hat{S}(t)\, e^{z_{0.975}\sqrt{\hat{Var}(\ln(\hat{S}(t)))}}\right]$$



one year survival rate? 60%
median surv. time? t*
(if we don't have 50%, we take the first moment we go ↓ 50%)

LOG-RANK TEST = test for comparing the survival distributions of two groups (two/more)

$$\begin{cases} H_0: S_1(\cdot) = S_2(\cdot) \\ H_1: S_1(\cdot) \neq S_2(\cdot) \end{cases}$$

: it observes the proportion of the rate of events over the time for each group, compares these observations with what should be expected if the groups would be the same and then it makes an assesment using χ^2 distr.

for K groups :

$H_0: S_1(\cdot) = \dots = S_k(\cdot)$

$H_1: \exists i,j : S_i(\cdot) \neq S_j(\cdot)$

HAZARD RATIO = ratio of the hazard rates corresponding to the conditions described by two levels of an exploratory variable

= chance of an event occuring in group 1 / chance of an event occuring in grop 2

$$HR = \frac{O_1/E_1}{O_2/E_2} = \begin{cases} = 1 \\ > 1 \\ < 1 \end{cases}$$

= 1 no effect

> 1 reduction in the hazard (= increase in survival) ⟹ group at numerator is a protective factor

< 1 increase in the hazard (= decrease in survival) ⟹ group at numerator is a risk factor

O_k = observed events

E_k = expected events

Ex.

	O_K	E_K
Female	59	75.9
Male	126	109.1

$$\Longrightarrow HR = \frac{59/75.9}{126/109.1} = 0.673$$

➔ The risk of deaths in females is 0.673 times the risk of deaths in males

➔ HR < 1 : females have higher survival probability than males

➔ Being a female is a protective factor

PROPORTIONAL-HAZARD COX MODEL → adding explanatory variables

$$\underbrace{h_i(t|X_i)}_{\text{hazard function for the i-th patient}} = h_0(t)\, e^{X_i^T \beta} = h_0(t)\, e^{\{\beta_1 x_{i1} + \dots + \beta_p x_{ip}\}}$$

$X_i \in \mathbb{R}^p$ = covariates of the i-th patient

⟶ ADJUSTED SURVIVAL CURVES :

$$\underbrace{S_i(t|X_i) = \left(S_0(t)\right)^{e^{X_i^T\beta}}}_{\text{adjusted survival function for the i-th patient}}, \quad S_0(t) = e^{-\int_0^t h_0(u)du}$$

cox.zph() → $\begin{cases} H_0: \text{hazard are proportional} \\ H_1: \text{hazard are } \underline{\text{not}} \text{ proportional} \end{cases}$